DOCUMENT RESUME

ED 330 604 SO 030 254

AUTHOR Dinkelman, Todd D.
TITLE Critical Thinking Content in Verbal Problems of the P-SAT and P-ACT+. Critical Thinking Reports Number 7.
INSTITUTION Illinois Univ., Urbana. Dept. of Educational Policy Studies.
PUB DATE 90
NOTE 62p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; Content Analysis; Content Validity; *Critical Thinking; Educational Research; High Schools; High School Students; Questionnaires; *Standardized Tests; Student Educational Objectives; *Test Content; Testing Problems; Testing Programs; Verbal Tests
IDENTIFIERS Preliminary American College Test Plus; *Preliminary Scholastic Aptitude Test

ABSTRACT

Education reformers have been concerned that U.S. schools are not helping to develop in their students critical thinking or reasonable, reflective thinking about what to do or believe. Because of the significant role standardized tests play in measuring student progress, and the way in which they may influence the content and dissemination of what is taught, it is important to know the extent to which standardized tests have critical thinking content. This paper investigates the amount of critical thinking content in the verbal questions of two widely used standardized examinations--the P-SAT and the P-ACT+. It was found that the P-ACT+, with 31.4% of its verbal based questions determined to contain critical thinking, scored a higher incidence of critical thinking content than did the P-SAT, with only 18.5% of its questions showing any critical thinking content. Whether standardized tests are compatible with the goal of promoting critical thinking is thus called into question. A 15-item list of references is included as is an evaluation questionnaire for determining critical thinking content in tests. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Illinois Critical Thinking Project

List of Reports

Bruce L. Stewart (1978), Teaching Logic in Grades 7-12: A Literature Review, (37 pages), Critical Thinking Reports, Number 1. ERIC document-- 183 587.

Bruce L. Stewart (1979), Testing for Critical Thinking: A Review of the Resources, (134 pages), Rational Thinking Reports, Number 2. ERIC document-- 183 588.

Thomas N. Tomko and Robert H. Ennis (1979), Evaluation of Informal Logic Competence, (53 pages), Rational Thinking Reports, Number 3. ERIC document-- 183 589.

Stephen P. Norris (1979), The Dependability of Observation Statements, (62 pages), Rational Thinking Reports, Number 4. ERIC document-- 183 590.

Michelle B. Fisher (1989), A Pilot Study on the Extent to which a Standardized Subject-Matter Test Assesses Critical Thinking Abilities, (17 pages), Critical Thinking Reports, Number 5. ERIC document-- in process, May, 1990.

Jim Kotowski (1990), Critical Thinking Content in the Iowa Tests of Educational Development, (33 pages), Critical Thinking Reports, Number 6. ERIC document-- in process, May, 1990.

Todd D. Dinkelman (1990), Critical Thinking Content in Verbal Problems of the P-SAT and P-ACT+, (57 pages), Critical Thinking Reports, Number 7. ERIC document-- in process, May, 1990.

CRITICAL THINKING CONTENT IN VERBAL PROBLEMS OF THE P-SAT AND P-ACT+

BY

TODD D. DINKELMAN

CRITICAL THINKING REPORTS

NUMBER 7

General Editor of Critical Thinking Reports
Robert H. Ennis

Published by

The Illinois Critical Thinking Project

Department of Educational Policy Studies
University of Illinois at Urbana-Champaign
Urbana, Illinois
1990

TABLE OF CONTENTS

Introduction ---- 1
Test Descriptions ---- 4
Procedure ---- 7
Results ---- 11
Discussion ---- 15
P-SAT ---- 17
P-ACT+ ---- 24
Other Considerations ---- 34
Conclusion ---- 45
References ---- 51
Appendix
Evaluation Questionnaire for Determining Critical Thinking Content in Tests ---- 53

Introduction

The 1980s witnessed a spate of literature describing the woeful status of schooling in the present-day United States. Much of this literature contains prescriptions for change. An early and influential report published by the National Commission on Excellence in Education (1983), A Nation at Risk, for example, decries the "rising tide of mediocrity" (p. 1) in which American school systems are sinking fast, leading to nothing less than "unthinking unilateral educational disarmament" (p. 1). Also advanced by the Commission is a plan to rescue the imperiled schools. Although not many of the subsequent reform reports quite match A Nation at Risk in its use of figurative language, and although there are varied assessments of the schools' problems, some common complaints are touched upon in many reports.

One such common concern is that American schools are not facilitating the development in their students of critical thinking -- defined in this study as reasonable, reflective thinking about what to do or believe. Boyer (1983), in High School, recognizes this problem and apparently is concerned by its presence. He lists four basic goals of his reform recommendations; the first of these states, "The high school should help all students develop the capacity to think critically and communicate effectively through a mastery of language"

(Boyer, 1983, p. 66). Another report, Horace's Compromise, by Sizer, addresses the problem of schools wherein students do not learn to think critically. Sizer notes of recent high school students,

> What is especially troubling is the low level of their reasoning skills, the abilities of analysis and synthesis. While students seem to be improving in rote-level, concrete learnings -- vocabulary recognition and, in mathematics, simple addition, for example -- their ability to think critically and resourcefully is lamentably weak and is continuing to weaken. (Sizer, 1984, p. 58)

Judging by these and other school reform reports of the last decade, critical thinking seems to be an essential aim for American schools. While there does exist considerable disagreement about how specifically to satisfy this aim, a common concern in many of the reports is that there is much improvement to be made in the teaching of critical thinking.

Along with critical thinking, a concurrent emphasis in many reform proposals is the need to hold schools accountable for the progress, or lack of progress, made in improving student performance. For a number of reasons, not the least of which is convenience, the standardized multiple-choice exam is the most prevalent method used to determine student performance and, in turn, school districts', individual schools',

and teachers' success as well. One consequence of using standardized tests to determine educational success or failure is the practice of "teaching to the test," wherein the subjects of lessons are dictated by the perceived subjects of the test the students will take. When critical thinking is valued as an aim of schooling and teachers are teaching to standardized tests, the question of the extent to which these standardized tests have critical thinking content is crucial. Niell and Medina note, "As teaching becomes 'coaching for the test' in many schools, real learning and real thinking are crowded out. Among the instructional casualties are the higher-order thinking skills" (Niell & Medina, 1989, p. 694). As this passage reflects, if the standardized tests used are not perceived to test for critical thinking abilities, they likely will promote pedagogical practices that are not conducive to the development of critical thinking abilities and thus compound the school's inability to meet its goal of facilitating critical thinking.

Two of the most widely used standardized tests in American high schools today are the P-SAT Test published by Educational Testing Service, and the P-ACT+ published by the American College Testing Program. In 1989, 298,159 students in 6,236 schools took the P-ACT+ exam. This test was used in every state in the United States. Only in Washington DC was the P-ACT+ not used (R. Clark, personal communication, April 6,

1990). In 1989, 1,490,860 high school sophomores and juniors took the P-SAT in all 50 states (B. Powell, personal communication, April 10, 1990). Also very widely used are the junior and senior level counterparts to these tests, the ACT and the SAT. Despite a decline in the number of 18-year-olds, in 1988, "1,134,364 students took the S.A.T.- a 5 percent increase over 1987 and a 13 percent jump since 1986. There were 842,322 students who took the A.C.T. this year -- 8.3 percent more than last year" (Walsh, 1988, p. 23).

Knowing the amount of critical thinking content a test contains is helpful in coming to understand the ways in which that test might influence curricular and pedagogical decisions. For this reason, and due to the widespread popularity of these tests and the potential consequences they have for so many students, the P-SAT and the P-ACT+ are worthy of investigation. Thus, this paper will investigate the amount of critical thinking content in the P-SAT and the P-ACT+ exams, and will include a discussion of the results of this investigation.

Test Descriptions

Both the P-SAT and the P-ACT+ are primarily designed as tests for high school sophomores and juniors. The SAT and/or the ACT tests are generaly taken by seniors, especially those who plan to go on to college after high school, although some students take the these two exams as juniors. Respectively, the

P-SAT and the P-ACT+ bear an extremely close resemblance, both in format and content, to the SAT and the ACT. The P-SAT and the P-ACT+ are both strictly multiple-choice and given in one sitting.

The P-SAT contains two sections: a verbal section and a mathematics section. The verbal section consists of 65 questions to be completed in 50 minutes. This section has four types of questions. The first 20 questions in the verbal section, antonym problems, posit a word and ask the test-taker to select the word, from a list of five words, that is most nearly opposite in meaning to the given word. The next ten questions, fill-in-the-blank problems, present sentences with one or two words omitted, and the test-taker is to select the word or pair of words from five words or word pairs that best fits the meaning of a given sentence when inserted in the sentence. The following 15 questions, relation problems, ask the student to choose one of five pairs of words that expresses a relationship similar to the relationship expressed by a given pair of words. The final set of questions, reading comprehension problems, the last 20 questions in the verbal section, ask the test-taker to read a passage of text and then to answer three to five questions based on its content. The math section consists of 50 questions to be completed in 50 minutes. The questions are drawn from the fields of arithmetic, algebra,

trigonometry, and geometry. Many of the mathematics questions are stated as word problems.

The P-ACT+ is divided into four sections: a Writing Skills Test, a Mathematics Test, a Reading Test, and a Science Reasoning Test. The Writing Skills Test has 50 questions to be answered in 30 minutes. This section consists of four textual passages that contain underlined parts that may or may not be grammatical errors and paragraphs that may or may not be in the most logical order. Mostly, students are asked to select the answer that corrects any grammatical error in accordance with the rules of standard written English or that is most consistent with tone or style of the passage as a whole. They are also asked to arrange the paragraphs in the passages in the best logical order. The mathematics test consists of 40 questions to be answered in 40 minutes. Like the P-SAT, these questions are drawn from the fields of arithmetic, algebra, trigonometry, and geometry; and many of the mathematics questions are stated as word problems. The Reading Test consists of 25 questions to be answered in 20 minutes. There are three passages and eight or nine questions for each passage based on the content of the passages. These questions are of roughly the same format as the reading comprehension problems found on the P-SAT. The final section of the P-ACT+, the Science Reasoning Test, contains 30 questions to be answered in 25

minutes. There are five passages in this section of the form examined in this study, two dealing with astronomy, two with plant biology, and one with chemistry. Test takers are asked to make inferences about results of experiments and about experiment construction.

Procedure

In determining the critical thinking content in the P-SAT and P-ACT+, the Evaluation Questionnaire for Determining Critical Thinking in Tests was used. This Questionnaire is a revision of Ennis' "Taxonomy of Critical Thinking Dispositions and Abilities" (Ennis, 1987, pp. 12-15). This taxonomy is the result of many years or refinement and is a well-respected conceptualization of critical thinking. The Questionnaire is in a question-based format thought to be clearer and more practical than the taxonomy for the purpose of evaluating tests. This Questionnaire, located in the appendix, contains the abilities and dispositions of critical thinkers based on the earlier stated working definition of critical thinking -- reasonable, reflective thinking about what to do or believe.

The tests were evaluated as follows: Scorer #1 administered both the P-SAT and the P-ACT+ to himself, scoring each question according to the ability tested for, if any, on the Questionnaire. Those questions not felt clearly evident by scorer #1 (46 out of the 215 question on both tests) were

administered to scorers #2 and #3 who evaluated the questions using the Questionnaire. Subsequently all three scorers compared their results. Whenever a test item was scored differently by one or more of the scorers, the item was discussed between them until there was mutual agreement.

In assigning the abilities to questions an attempt was made to apply the categories in the Questionnaire as specifically as reason permitted. It should be understood that most abilities listed under the Clarification and Basic Support sections, numbers 1 through 8 on the Questionnaire, require abilities listed under the Inference section. When an item was judged to test only for an ability listed under the Clarification or Basic Support sections, only that ability was recorded. In such cases, an Inference ability was not also assigned. The only items assigned both a Clarification or Basic Support ability and an Inference ability were those items judged to test for an ability under the Clarification or Basic Support sections and a separate and distinct ability under the Inference section. For an example of such a question, see Example 11 (p. 32).

The assignment of a critical thinking ability to a question in each instance was solely an analysis of critical thinking content. Thus, the results of these assignments should be viewed as judgments of critical thinking content validity. That is, the scorers sought to determine for what critical thinking

content the test items tested. "In its perennial form," reports Messick, "content validity is based on expert judgments about the relevance of the the test content to the content of a particular behavioral domain of interest and about the representativeness with which item or task content covers that domain" (Messick, 1989, p. 7). In this case, the domain of interest is critical thinking. No attempt was made to determine the cognitive processes a test-taker might actually use in answering the test item, a construct validity issue. Although issues of construct validity have vitally important ramifications for educational theory and policy, a construct validity analysis of these tests is a project beyond the scope of this study, in part because of the resources required for such a study, and more importantly, there are a number of compelling reasons for undertaking a content analysis. First, in instances of test-driven teaching and curriculum development, it is the actual content of the test upon which teachers, administrators, and curriculum specialists, make curricular decisions. Thus for those who would like to see critical thinking as an educational objective, it becomes important to investigate the content of these tests. Second, a study of critical thinking content may emphasize the importance of critical thinking content to future test writers. And in addition to the potential for practical improvements in critical thinking instruction and testing, such

a content analysis may serve as a stimulus to further research into content and construct validity in critical thinking.

The analysis of the tests in this study is an analysis of the verbal-based questions on the tests. This analysis excluded mathematic reasoning, and thus excluded the mathematics section of both tests. Although this approach limits the scope of the study, the current level of conceptualization about critical thinking in mathematics has not reached the point where critical thinking abilities could be assigned with an adequate degree of certainty. As Berg notes, in judging mathematical items for critical thinking content one must make assumptions about the level of math ability of the students taking the test, relying on an extensive knowledge of mathematics and its development in school curricula, a difficult task for two tests given to so many students in so many different school districts. For non-mathematical items on general educational achievement exams like the P-SAT and the P-ACT+, the problems associated with making these assumptions are lessened (Berg, 1990, pp. 6-10). Furthermore, the dominance of verbal reasoning used in making decisions in both the private and public realms indicates the importance of assessing the verbal-based questions on these tests for critical thinking content.

As well as research needed on the relationship of mathematics and critical thinking, further work is also needed in assessment methods for critical thinking in tests. This researcher is unaware of any reliable and widely accepted procedure for this purpose. Thus the procedure undertaken in this study is exploratory and subject to further revision, as is the Evaluation Questionnaire for Determining Critical Thinking in Tests. Although the main task of this study is to make a statement about the critical thinking content of two widely used multiple-choice exams, an additional objective is to assess the strengths of this method as an evaluative procedure. In this sense, it is hoped that this project will contribute to the production of a reliable procedure enabling future analyses of tests for critical thinking content.

Results

Table 1

Critical Thinking Content in P-SAT

Section 1

Verbal

Section	CT Ability
1-45	none
46	Induct.--best expl.--interp. of authors intended meaning

Table 1 (continued)

47	none
48	Induct.--best expl.--interp. of authors intended meaning
49	Induct.--best expl.
50	none
51	none
52	Induct.--best expl.
53	Induct.--best expl.
54	Induct.--best expl.
55	none
56	Induct.--best expl.
57	Induct.--best expl.--interp. of authors intended meaning
58	Induct.--best expl.
59	none
60	Arg. analysis; Induct.--best expl.
61	Deduction
62	none
63	Induct.--best expl.
64	none
65	none

Section 2

Mathematics

Section	CT Ability
1-50	not evaluated

Table 1 (continued)

Note. Total # of verbal P-SAT questions: 65.

Percent of verbal questions testing critical thinking content: 18.5%.

Percent of abilities and dispositions evaluated: 12.5% (3/24).

Table 2

Critical Thinking Content in P-ACT+

1- Writing Skills Test	CT Ability
1-24	none
25	Induct.--best expl.--interp. of authors intended meaning
26-37	none
38	Induct.--best expl.
39-50	none
2-Mathematics Test	**CT Ability**
1-40	not evaluated
3- Reading Test	**CT Ability**
1	Induct.--best expl.
2	Induct.--best expl.
3	none
4	none
5	Induct.--best expl.
6	Induct.--best expl.
7-16	none
17	Observation eval.--inference

Table 2 (continued)

18	Induct.--best expl.
19	Arg. analysis--identify concl.
20-25	none

4- Science Reasoning

Test	CT Ability
1	none
2	none
3	Induct.--best expl.
4	Induct.--generalization
5	Induct.--generalization
6	Induct.--best expl.--experiment design
7	Deduction; Induct.--best expl.
8	Induct.--best expl.
9	none
10	none
11	Induct.--best expl.--experiment design
12	Induct.--best expl.
13	none
14	none
15	Deduction; Induct.--best expl.
16	Deduction; Iduct.--generalization; Induct.--best expl.
17	Deduction; Induct.--best expl.
18	Induct.--best expl.
19	Induct.--best expl.--causal claim
20	Deduction; Induct.--best expl.--causal claim

Table 2 (continued)

21	Induct.--best expl.--experiment design
22	Deduction; Induct.--best expl.
23	Induct.--best expl.--causal claim
24	Induct.--best expl.--experiment design
25	Arg. analysis--ident. conclusions; Deduction
26	Arg. analysis--ident. conclusions; Arg. analysis--ident. unstated reasons; Deduction
27	Deduction; Arg. analysis--ident.--unstated reasons
28	Induct.--best expl.--seek evidence
29	Arg. analysis--ident. conclusions; Induct.--best expl.
30	Arg. analysis--ident. conclusions; Induct.--best expl.

Note. Total # of verbal P-ACT+ questions: 105.

Percent of verbal questions testing critical thinking content: 31.4%.

Percent of abilities and dispositions represented: 20.8% (5/24).

Discussion

The results are discouraging to one who thinks it is important that critical thinking be measured on achievement tests that are used to indicate a student's overall educational success in high school. Simply put, there is not a great deal of critical thinking content on the analyzed portions of these tests,

tests designed to aid beginning high school students in plotting their course of study. Of the combined total of 170 verbal questions evaluated on both tests, only 45, or 26.5%, were judged as testing for critical thinking content. The P-SAT was found to test critical thinking on 18.5%, or 12 of its 65, verbal questions. The P-ACT+ fared better, as 33 of its 105 verbal questions, or 31.4%, were scored as questions employing critical thinking content. In both exams, induction by inference to best explanation was far and away the most frequently found ability or disposition. In the P-SAT, three different abilities were present, but induction by inference to best explanation accounted for 84.6% of those questions judged as testing for critical thinking content. In the P-ACT+, five different abilities were present; and induction by inference to best explanation accounted for 56.5% of the critical thinking content questions, deduction accounted for 19.6%, and argument analysis accounted for 15.2%. A more detailed discussion of the results follows.

Overall, for advocates of critical thinking in the schools, a rate of less than one in five verbal questions for the P-SAT and less than one in three verbal questions for the P-ACT+ might seem rather unimpressive. On the other hand, it should be noted that neither the P-SAT nor the P-ACT+ is advertised as a critical thinking test, but rather as a general assessment of

educational development. Perhaps these figures will prompt discussion about what the ideal would be for general tests of educational development.

The following discussion will examine more closely what was found in each test. In particular, the specific critical thinking abilities tested will be discussed, and examples of these, as well as non-examples, will be given. This will be followed by a discussion of these following considerations: (a) the breadth of coverage on both exams of the abilities and dispositions found in the Evaluation Questionnaire for Determining Critical Thinking in Tests, (b) the limitations imposed on evaluating critical thinking content by the multiple-choice format, (c) the sparsity of critical thinking content in non-science sections of the exams as compared to the P-ACT+'s Science Reasoning section, (d) the amount of subject-specific knowledge familiarity required in order to use a critical thinking ability on some questions, (e) the effectiveness of the procedure used in evaluating these exams for critical thinking content.

P-SAT

Antonym problems. Of the four types of questions in the verbal section of the P-SAT, only the reading comprehension problems were found to test for critical thinking content. Thus the first 45 questions on the exam are void of critical thinking

content. The first 20 of these questions, antonym problems, posit a word and ask the test-taker to select the word, from a list of five, that is most nearly opposite in meaning to the given word. The following example is typical of this type of question:

Example 1. DOWNCAST: (A) spiteful (B) cheerful
(C) innocent (D) surprised (E) refined

(P-SAT, 1988, p. 2)

This type of question seems merely to be testing for knowledge of word meanings and not any ability associated with reasonable and reflective decision making, and therefore fails to qualify as testing for critical thinking content.

Fill-in-the-blank promlems. The next ten questions, fill-in-the-blank problems, also judged void of critical thinking content, present sentences with one or two words omitted, and the test-taker is to select the word or pair of words from five words or word pairs that best fits the meaning of the sentence when inserted in the sentence. The following example is typical of these questions:

Example 2. Because the professor's theories were so intimately linked with his private life, he interpreted a criticism of the former as an attack on the latter as well; hence, ---- disagreements became ---- insults.

(A) basic . . metaphysical

(B) rancorous . . unintelligible

(C) innocuous . . deserved

(D) conceptual . . personal

(E) seminal . . intellectual (P-SAT, 1988, p. 3)

The correct answer is (D). Again, this type of question appears to be testing for knowledge of word meanings and not for any critical thinking content.

Relation problems. The relation problems, the following 15 questions, ask the student to choose one of five pairs of words that expresses a relationship similar to the relationship expressed by a given pair of words. As with the previous two types of question, the relevant domain for a judgment of this question's content validity seems to be knowledge of the meanings of words in the English language, not critical thinking. The following question, correctly answered (E), is representative of this group of questions.

Example 3. DROWSINESS : SLUMBER ::

(A) embarrassment : threaten (B) poverty : share

(C) sogginess : drink (D) assurance : fret (E)

hunger : dine (P-SAT, 1988, p. 4)

So it is only in the final 20 questions in the verbal section that all of the critical thinking content questions were found.

Reading comprehension problems. Of these 20 questions, 12, or 60%, were judged as employing critical thinking content. These final 20 questions all followed the format wherein the

test-taker is asked to read a passage of text and then to answer three to five questions based on the content of the passage. Five passages were presented. The most often found critical thinking ability was the ability to induce by inference to best explanation. 11 of the 12 critical thinking content questions were judged to employ this ability. These results indicate an overreliance on one ability and the absence of other abilities entirely. An example of a question judged as induction by inference to best explanation is the following:

Example 4. The passage as a whole is best characterized as

(A) an analysis of a Black soldier's political beliefs

(B) a description of a historical person

(C) a reexamination of the early history of Stockbridge

(D) an evocation of the challenges and rewards of an earlier era

(E) a reevaluation of common assumptions about the American Revolution (P-SAT, 1988, p. 6)

The correct answer is (B). Although the correct answer is not an explanation, but rather a characterization, the process of selecting the best characterization of the passage as a whole from a list of alternatives is much like the inductive move of selecting from a list of alternative explanations of why the passage was written. Thus this question was scored as induction to best explanation. An additional reason why this passage qualifies as a question testing for induction to best

explanation is that the evidence provided by the text must be weighed in light of each proposed answer. Not just knowledge of the meanings of isolated words is required, but the "passage as a whole" must be taken into account, and a judgment about the best characterization is required. In other words, the answer is not self-evident due to one's knowledge of terms such as "analysis," "description," "reexamination," "evocation," and "reevaluation," although knowledge of the meaning of these terms is necessary. Reflection is called for, and therefore the knowledge alone is not sufficient to determine the correct answer. In contrast to Example 4., the set of questions based on the same passage contains the following question that did not qualify as having any critical thinking content:

Example 5. The passage suggests that Hull was all of the following during his life EXCEPT

(A) medical assistant

(B) army pensioner

(C) local government official

(D) small-scale farmer

(E) husband and father (P-SAT, 1988, p. 6)

In this case, the correct answer (C) requires that one simply locate in the passage all of the other possibilities or recognize the absence of any reference to Hull as a "local government official." At its most this question seems to be testing for

exemplification, the ability to name/recognize a given example as a member of a conceptual category or to name/recognize examples of a given conceptual category. Although exemplification is an important and often used ability, it is not recognized as a critical thinking ability because it calls only for knowledge of specific conceptual categories and the recognition of members of that category. In this sense, exemplification is not reasonable, reflective thinking. The question does not ask the student to make a qualitative judgment about the best possible answer as was the case in Example 4.

Critical thinking abilities other than induction to best explanation. In the reading comprehension problems, only two other abilities were tested in addition to induction to best explanation: "analyze the structure of arguments" was found in one question along with "induce by inference to best explanation," and "deduce" was found in one question. Thus, the one critical thinking content question that was judged as a non-induction to best explanation question was judged as a deduction question. The part of the passage that seems to be especially relevant to this question and the question itself are as follows:

Example 6. "As a general summary we can say that the left hemisphere is the seat of most of the capacity for language, calculation, and logical analysis. The right

hemisphere is better than the left at spatial perception, recognition of faces, and musical appreciation."

It can be inferred from the passage that the right hemisphere of the brain would be more useful than the left in performing which of the following tasks?

(A) Learning a foreign language

(B) Studying for a mathematics examination

(C) Recognizing an operatic theme

(D) Analyzing a problem

(E) Figuring income tax (P-SAT, 1988, p. 8)

The correct answer is (C). This example qualifies as a deduction item because the conclusion necessarily follows from the premises. If the premises are a) "the right hemisphere is better than the left at spatial perception, recognition of faces, and musical appreciation" and that b) "better" can be interpreted to mean "more useful," then by deduction it follows that "Recognizing an operatic theme" is the correct answer. It is important to note here that although this item qualifies as testing for deduction, this item is arguably not a very good test of one's ability at deduction. This is such an elementary case of deduction that the answer nearly appears to be self-evident. Scoring this question as testing for deduction is rather generous because the item seems to rely heavily on exemplification. That is, the item seems to rely heavily on seeing that

"recognizing an operatic theme" is an instance of musical appreciation. However, it was felt that there was warrant to label this as deduction, especially when compared to other exemplification questions on the test. Thus it should be kept in mind that the purpose of this report is not to make claims about the sophistication of the questions judged as having critical thinking content validity, although a question's sophistication is an important criterion to consider when determining the strength of a test as an assessment instrument for critical thinking.

P-ACT+

The P-ACT+ fared better than the P-SAT as a test employing critical thinking content. Of the 105 verbal questions surveyed, 33 were judged as testing for at least one critical thinking ability, or 31.4%, as compared to the P-SAT's 18.5%. Further, nine of the 105 verbal questions tested for two or more critical thinking abilities, or 8.6% as opposed to 1.5% for the P-SAT. Yet the overall percentages give one an incomplete picture of the strength of the P-ACT+ as a test for critical thinking abilities. The breakdown of percentages by individual tests that make up the P-ACT+ contributes additional clarity. The Writing Skills Test has 50 questions of which only two, or 4%, contain critical thinking content. Of the Reading Skills Test's 25 questions, 7, or 28%, contain critical thinking

content. The Science Reasoning Test appears to live up to its billing as a reasoning test by registering 24 of its 30 questions, or 80%, as valid for critical thinking content. So the P-ACT+ distinguishes itself from the P-SAT in terms of critical thinking content largely on the basis of the presence of the Science Reasoning Test. Excluding this part of the exam, the P-ACT+ records only 9 of 75 verbal questions (the Reading Skills Test and the Writing Skills Test combined), or 12%, as critical thinking content items, while the P-SAT recorded 12 of 65 or 18.5%.

Writing Skills Test. The Writing Skills Test (Part 1) consists of 50 questions based on four passages with underlined text upon which the questions are based. According to the publishers of the P-ACT+, the American College Testing Program (ACT), the Writing Skills Test "stresses the analysis of the kinds of prose that students are required to read and write in many secondary and postsecondary curricula, rather than the rote recall of rules of grammar" (P-ACT Plus, 1988, p. 4). Of what is meant by 'analysis' here, nothing is stated. By this study, the 'analysis' to which ACT refers could not be called critical thinking analysis because only 4% of the questions in this section call for critical thinking. Furthermore, the stress of many questions seemed to belie the claim of not stressing "the

rote recall of rules of grammar." The following is a typical example of the form and content of these questions:

Example 7. ". . . It is seldom observed, however, that there <u>is</u> striking similarities between the realms of sports and music. . ."

A. NO CHANGE

B. certainly is

D. are

E. is certain (P-ACT+, 1987, p. 2)

The task is to choose from the list the best alternative to the underlined piece of text. The correct answer, D., does not demonstrate critical thinking content, but rather seems to be testing for knowledge of standard written English, more specifically the application of the correct form of the verb 'to be.'

Only two of the questions on this part of the exam tested for any critical thinking ability, and both of these tested for induction to best explanation. The following question is one of the questions judged as "induce by inference to best explanation:"

Example 8. This passage was probably written for readers who:

A. are experts in Arthurian literature

B. are authorities on ancient British civilization and culture.

C. are convinced that Arthur was an actual historical personage.

D. have some familiarity with the Arthur legends.

(P-ACT+, 1987, p. 5)

The correct answer is D. The defense of this question as one testing for induction to best explanation is similar to the defense given for Example 4. In both examples, the evidence provided by the text must be weighed in light of each conclusion in selecting the best answer. More than just knowledge of the meanings of isolated words and the rules of standard written English are called for in this question; the entire passage must be taken into account, and a judgment about the best explanation of the author's intended audience is required. Also, although the correct answer is not in the form of an explanation, this question can be seen as testing for induction to best explanation because it seems to ask that the student weigh alternative explanations for why one should conclude that this passage was written for each of the four possible audiences.

Reading Test. Of the Reading Test, ACT reports, "The test focuses upon the complex of complementary and mutually supportive skills that readers must bring to bear in studying

written materials across a range of subject areas, rather than the rote recall of facts, information from outside the passages, isolated vocabulary items, or rules of formal logic" (P-ACT Plus, 1988, p. 4). The skills that make up "the complex" referred to here are not specified by ACT in this source. According to the results of this study, only three critical thinking abilities are included. Five of the seven questions judged as critical thinking content questions on the Reading Test are induction to best explanation questions. Of the other two, one question calls for argument analysis, and the other calls for evaluation of observation credibility. For the sake of illustrating how questions were judged according to the Questionnaire, and because there have been several examples of induction to best explanation questions already, these two items that are not induction to best explanation questions follow:

Example 9. On the basis of the information in the passage, which of the following statements is most clearly an opinion rather than a fact?

A. The English language contains examples of Greek influence.

B. The Romans conquered the Greeks after the Macedonians did.

C. The Greeks used the ideas of other civilizations to develop mathematical concepts.

D. Adoption of Plato's concept of government would have prevented the decline and conquest of Greece.

(P-ACT+, 1987, p. 27)

This was judged as an evaluation of the ability to "judge the credibility of an observation--amount of inference involved." In assigning this ability to this question it was assumed that the test-makers used the terms 'fact' and 'opinion' in the sense that a 'fact' is a description of something requiring no inference and that an 'opinion' is a description of something requiring some inference (Weddle, 1985, pp. 19-26). Again it should be stressed that the purpose of this report is not to make claims about the sophistication of the questions judged as having critical thinking content validity. As with Example 6., although this question qualifies as having content validity for critical thinking, it is an elementary instance of a critical thinking ability. The answer is nearly self-evident to one who is familiar with the popular distinction between fact and opinion. Many people might be able to answer this question without reflecting about the passage and the potential answers to the question. Again, in evaluating the strength of these tests one must consider not only the number of questions that test for

critical thinking content, but judgments about the sophistication of the questions must also be considered.

The other critical thinking content item that was not judged an induction to best explanation question in the Reading Test was judged as "argument analysis--identify conclusions."

Example 10. The main idea of the passage is that:

A. reading ancient Greek myths today might cause some people to change their religious beliefs.

B. the ancient Greeks have had considerable influence on the development of Western civilization.

C. the ancient Greeks answered most questions concerning nature.

D. many philosophical statements are to be found in Greek literature. (P-ACT+, 1987, p. 26)

The correct answer is B. In this case, the task of deciding on the main idea of the passage was judged similar enough to the task of identifying the conclusion of an argument to warrant the assignment of "argument analysis--identify conclusions" to the question. Consistent with the method described in the procedure section of this paper, only "argument analysis--identify conclusions" is recorded in the results for this question even though any instance of argument analysis makes use of induction to best explanation. Again, only items that seemed to

ask specifically for separate abilities were scored as containing more than one ability. For an example of such a question, see Example 11.

Science Reasoning Test. The Science Reasoning Test scored the highest incidence and the widest range of critical thinking abilities. 80% of its questions were found to test for at least one critical thinking ability, far surpassing the percentage of critical thinking content items in any other sub-test on either test. According to ACT, "The items require students to recognize and understand the basic features of, and concepts related to, the provided information; to critically examine the relationships between the information provided and the conclusions drawn or hypotheses developed; and to generalize from given information to gain new information, draw conclusions, or make predictions" (P-ACT Plus, 1988, p.4). Certainly some of these skills pointed to by ACT, such as critically examining relationships, generalizing, and making predictions, can be seen as critical thinking abilities. Thus ACT is arguably making a claim that the Science Reasoning Test tests for critical thinking abilities. The results of this investigation support such an interpretation.

The following table summarizes the findings:

Table 3

Critical Thinking Abilities in Science Reasoning Test

Critical Thinking Ability	#of items	% of total in Science Reasoning Test
Induct.--best explan.	11	36.7
--causal claims	3	10.0
--experiment design	4	13.3
--seek evidence	1	3.3
Induct.--generalization	4	13.3
Deduction	8	26.7
Arg. analysis--ident. conclusions	4	13.3
Arg. analysis--ident. unstated reasons	2	6.7

Note. The more specific "Induct.--best explan." items (i.e. causal claims, experiment design, and seek evidence) are not included in the number or percent for "Induct.--best explan."

Since examples of induction to best explanation, argument analysis, and deduction have already been included in this discussion, and since all of the previous examples were judged as containing only one critical thinking ability, the following example from the Science Reasoning Test is a

problem that tested for more than one ability. The final five questions in this test are based on a passage which poses the question, "How did Earth's oceans originate?" Two differing views, those of Scientist 1 and Scientist 2, are presented and the test-taker is asked a series of questions regarding these two views. The last question of this series follows:

Example 11. The hypothesis of Scientist 1 would be supported by any evidence suggesting that 2 billion years ago Earth had:

F. less water in the oceans.

G. a smaller rate of cometary infall.

H. more water vapor in its atmosphere.

J. more ice in the polar regions. (P-ACT+, 1987, p.36)

The correct answer is F. This question was judged as containing both "argument analysis--identify conclusions" and "induction- best explanation." It contains "argument analysis--identify conclusions" because the test-taker must decide what the hypothesis of Scientist 1 is; the hypothesis is not explicitly identified in the passage attributed to Scientist 1. And the question also contains "induction--best explanation" because, once the hypothesis is identified, the test-taker must select the best answer to complete an explanation of why certain evidence would support Scientist 1's hypothesis.

Other Considerations

Breadth of coverage. So far in this discussion much attention has been paid to the presence of critical thinking content in the verbal questions on these exams. Percentages of the number of verbal questions containing critical thinking content have been given, but these figures give an incomplete picture of the exams' critical thinking content. The number of different abilities and dispositions is further revealing of this content. One can conceive of an instance such that, using this study's procedure, a multiple-choice test could register every question as testing for critical thinking content, but only testing for one ability, for example, deduction. In fact, there is such a test -- "Logical Reasoning" (Hertzka & Guilford, 1955). Presumably, on tests of general educational development, testing for just one ability to the exclusion of all other critical thinking content would be unacceptable. Fortunately, this was not the case for either the P-SAT or the P-ACT+.

Although neither exam tested for just one ability to the exclusion of all other critical thinking content, both tests covered only a narrow portion of the abilities and dispositions identified in the Questionnaire. In determining the following percentages, the Questionnaire is taken to contain 24 categories composed of12 major abilities, numbers 1-12 on the Questionnaire, and of12 major dispositions, numbers 13-24.

The P-SAT tested for only 3 of the 24 categories, or 12.5%, of those listed on the Questionnaire. The P-ACT+ fared slightly better testing for 5 of the 24 categories, or 20.8%. In these calculations, any ability more precisely determined than one of the major categories on the Questionnaire is counted as an instance of that major category; so, for example, question 21 on the Science Reasoning Test, which was recorded in the results as "Induct.--best expl.--experiment design," was counted as an instance of the major category induction by inference to best explanation, number 11, because "Induct.--best expl.--experiment design" is taken to be a special instance of "Induct.--best expl." In interpreting these figures, the contention, which will be made later in this discussion, that it may be impossible to test for dispositions on multiple-choice exams should be borne in mind. If this contention is true, then the highest figure for coverage one could expect on the P-SAT and the P-ACT+ computed by this method would be 50%. Whatever the case, omitting as many critical thinking content abilities as done on these tests seems questionable for tests designed as general educational achievement assessment tools.

Another consideration related to coverage of critical thinking content besides the number of different abilities and dispositions tested is the pattern of distribution of critical thinking abilities across the verbal questions judged as testing

for critical thinking content. In both exams, induction by inference to best explanation was far and away the most frequently found. In the P-SAT, three different abilities were present, but induction by inference to best explanation accounted for 84.6% of those questions judged as testing for critical thinking content. In the P-ACT+, five different abilities were present; and induction by inference to best explanation accounted for 56.5% of the critical thinking content questions, deduction accounted for 19.6%, and argument analysis accounted for 15.2%. It is possible that induction by inference to best explanation is more valuable or more important than any of the other critical thinking abilities; and, if this were the case, then there would be some justification for such a heavy emphasis on that ability. However, the lack of attention paid to the other critical thinking abilities may serve to undervalue these important abilities.

Multiple-choice and critical thinking. For those who advocate critical thinking as an educational ideal, the strict use of a multiple-choice format in these tests should be a source of concern. Multiple-choice exams have obvious economic advantages that make them more appealing than some other types of assessment tools. Multiple-choice exams can be administered to many test-takers simultaneously and relatively inexpensively, and they can be machine graded

quickly. Also, given an answer key, the results of the test can be consistently reproduced across graders, thus removing differences in grading based on the subjectivities of those grading.

But beyond these advantages, multiple-choice exams have drawbacks as well, especially for the purpose of evaluating critical thinking. One disadvantage is that, by emphasizing the one correct or best answer, these tests may give the impression that most problems have just one correct or best solution (Norris & Ennis, 1989, p. 29). Two of the dispositions on the list of dispositions used in this study are to look for alternatives and to be open-minded. The stress on selecting the one best or correct answer from a limited number of alternatives seems to run counter to these dispositions. A further and related disadvantage is that multiple-choice exams temper the role of creativity in critical thinking. With a multiple-choice question, the correct or best answer is always provided for the test-taker, along with the incorrect answers. The test-takers always decide between answers that are already provided and are never asked to create their own answers. Although not all critical thinking is creative thinking, there is a significant amount of good thinking that can be classified as both critical and creative. Multiple-choice tests do not emphasize thinking that is both critical and creative.

Multiple-choice exams can be further seen deficient as critical thinking assessment tools because they do not give a good indication of the test-takers' disposition to use their critical thinking abilities (Kotowski, 1990; Norris & Ennis, 1989). The conception of critical thinking upon which this study is based lists twelve dispositions of an ideal critical thinker, none of which was judged as present in either the P-SAT or the P-ACT+. Arguably, it may be impossible to test for dispositions in tests of this sort . In order to assess critical thinking dispositions, persons must be observed in situations where they have autonomy to make decisions about how to act and think. Multiple-choice exams do not provide such situations. For example, the structure of the multiple-choice exams commands the test-taker to heed the disposition listed on the Questionnaire "to take a position when the evidence and reasons are sufficient to do so;" but by doing so test-takers are not demonstrating that this is something they are disposed to doing under different and varied conditions. Judging critical thinking dispositions requires observations of persons in different and varied contexts that allow them to exercise autonomy.

Science Reasoning Test dominance. Another source of concern for those who advocate critical thinking as an educational ideal arises from the abundance of critical thinking

content in the Science Reasoning Test of the P-ACT+ and, by comparison, the scarcity of critical thinking content in all the other tests analyzed. 80% of the Science Reasoning Test's questions contained critical thinking content. The percentages for the other verbal sections of both tests were Reading Skills Test -- 28%, Writing Skills Test -- 4%, and the P-SAT's verbal section -- 18.5%. Such a high number for the Science Reasoning Test is a welcome sign to advocates of critical thinking in schools. While the addition of the Science Reasoning Test can be seen as a positive move by ACT, a move other developers of general educational achievement test should be encouraged to follow, there is concern that this uneven distribution over the three tests of critical thinking content might be perceived as a message that critical thinking is important in science but not important in non-science fields. One can reasonably assert that critical thinking is as important in literary criticism, rhetoric, and the social sciences for example, as it is in biology, chemistry, and physics. If this assertion is correct, then a strong argument can be made that there should be a more even distribution of critical thinking content across the sections of these exams, especially given the contention that the content of these exams is contributing to the shaping of curricula.

Subject knowledge requirements. Knowledge of the critical thinking content validity of an exam is helpful in

judging that exam's usefulness as an assessment of the important educational ideal of critical thinking, but such knowledge is not the only consideration in making this judgment. As mentioned earlier, one should also consider the difficulty of the items judged as testing for critical thinking. And still a third consideration is how much subject knowledge is required to answer the question. A prepnderance of test items may have a high presence of critical thinking content of a sophisticated variety, but may require detailed knowledge of a certain subject's terminology and methodology in order to answer the questions correctly. In this case, even a highly skilled and well practiced critical thinker might score poorly on this exam due to lack of familiarity with the technical language used in the exam questions.

Ideally, on an exam purported to test for critical thinking, the subject knowledge requirements simply would be minimized to the point that any test-taker with a basic understanding of the language in which the test is written would be well enough equipped to answer all questions without becoming tripped up by not knowing some piece of information not provided in the test. However, on tests of general educational achievement like the P-SAT and the P-ACT+ that claim to assess achievement in a variety of areas, not just critical thinking, the ideal amount of subject knowledge

requirements becomes much less clear. Whatever the ideal may be, it should be noted that many of the verbal questions analyzed in this report as containing critical thinking content required some subject knowledge in order to be answered correctly. This awareness is noted not to suggest that these requirements be seen as problems. The purpose of raising this issue is not to register complaints against the P-SAT and the P-ACT+ on this account, but rather to encourage discussion among those interested in testing for critical thinking about the ideal amount of subject knowledge requirements on general educational achievement tests.

The subject knowledge requirements seemed to be most stringent in the P-ACT+'s Science Reasoning Test, the test registering more critical thinking content than all the other analyzed verbal questions from both tests combined. For example, the following question from the Science Reasoning Test was scored as "Deduction" and "Argument Analysis--identify unstated reasons":

Example 12. According to the information presented by Scientist 2, if the cometary origin hypothesis for Earth's oceans is correct, which of the following planets is LEAST likely to have large amounts of water derived from comets?

A. Jupiter

B. Mercury

C. Saturn

D. Neptune (P-ACT+, 1987, p. 36)

The correct answer is B. This question was judged as testing for two abilities: "argument analysis--identifying unstated reasons" because one must infer from Scientist 2's argument the unstated reason that "planets closer to the sun are less likely to have water from cometary infall than planets farther from the sun," (p. 34) and deduction because this statement commits one to selecting the planet Mercury. However, test-takers not knowing the order of planets in the solar system could not deduce the answer. Even the most able critical thinkers could not answer this question correctly, except by guessing, if they did not know which of these planets is closest to the sun. Additionally one might argue that phrases such as "Increased CO2 concentration" (p. 34) and ". . . the effecto of NaHCO3 on the photosysnthetic rate of Elodea. . ." (p. 35) will sidetrack those unfamiliar with science not because of the critical thinking content but because of the technical language. Thus, the background knowledge requirements are an additional consideration in making decisions about an exam's usefulness for assessing critical thinking.

Evaluation of procedure. An auxiliary reason for conducting this study, besides the primary reason of assessing

the critical thinking content in the verbal questions of two widely used tests, was to contribute to the production of a reliable procedure enabling future analyses of tests for critical thinking content. The procedure used was exploratory and is subject to further revision. This study itself, in fact, suggests several considerations for improving this method. One weakness of this procedure is that no data about its inter-rater reliability are available. With the goal of producing a standardized scoring device useful for determining critical thinking content, inter-rater reliability becomes an important concern. Another, related, methodological concern is the procedure for reaching consensus among the three analysts when there were items about which they disagreed. Although each analyst appeared to apply the critical thinking abilities and dispositions in the discussion about disputed items, the ways that individual personality factors affect such discussions may influence the reaching of a consensus. The impact of personality factors in the reaching of a consensus could be minimized by carefully redesigning the procedure.

Another concern is the ease with which the Questionnaire can be applied to multiple-choice exams. There is room for modification of the procedure along these lines. For example, as mentioned earlier, assigning to a multiple-choice item any of the critical thinking dispositions seems only remotely, if at all,

possible. Also, the ease with which the Questionnaire can be used appears to depend heavily on the analyst's knowledge of critical thinking. In this study, all three analysts had a good deal of expertise in critical thinking, and their expertise was relied on in the discussions about disputed items. To make this procedure useful for those analysts who are unfamiliar with the field of critical thinking would require supplemental information regarding the use of this method, such as examples of questions analyzed as testing for specific abilities, definitions of critical thinking terms, and explanations of the Questionnaire's limitations.

One final consideration important in the evaluation of the procedure used in this study is its capacity to produce meaningful results. Using the method described in the Procedure section of this paper did produce meaningful results, but, due to the exploratory nature of this procedure, great care must be taken in making conclusions and generalizations based on the results. The results are meaningful in suggesting what critical thinking abilities the verbal questions seemed to be testing for on the P-SAT and the P-ACT+. The results should be considered as one of several important types of information to be considered when making judgments about an exam's usefulness as a critical thinking assessment device.

Conclusion

The purpose of this study was to determine the presence of critical thinking content in the verbal based questions of the P-SAT and the P-ACT+ exams, two exams used extensively in the United States as general tests of educational achievement for high school sophomores and juniors. The importance of these exams is reflected in the numbers of students who take them annually and the manner in which they are used in charting high school academic programs. Furthermore the importance of such knowledge is even more apparent given their near identity to their junior and senior level counterparts, the SAT and ACT exams. In light of the calls for holding teachers and schools more accountable for achievement in education that, in turn, encourage the practice of teachers' teaching to the test, which then promotes test-driven curricula, determining the critical thinking content of widely used tests becomes crucially important for those concerned with curriculum development.

The P-ACT+, with 31.4% of its verbal based questions assessed as containing critical thinking, scored a higher incidence of critical thinking content than did the P-SAT with only 18.5% of its questions. For advocates of critical thinking instruction in the schools such figures might seem rather low for tests of general educational achievement, and could serve

as starting points for discussion about what the ideal figures for such tests might be. Certainly such a discussion would be incomplete without taking into account the sophistication of the questions judged as testing for critical thinking content. Many of the evaluated questions on these two exams were elementary cases of critical thinking abilities. Thus one must be cautious not to infer conclusions giving these exams too much credit as tests of critical thinking abilities based solely on critical thinking abilities incidences of less than one in three verbal items for the P-ACT+ and less than one in five verbal items for the P-SAT.

In addition to the overall percentage scores for these tests, several other important considerations were noted in this report. First, the breadth of coverage on both exams was found to be limited. Although the P-ACT+ fared better on this account than did the P-SAT, neither tested for many of the abilities and dispositions found on the Questionnaire; and further, induction by inference to best explanation was found in far more questions than any other ability. Second, the drawbacks of multiple-choice exams in critical thinking assessment were noted. Potential disadvantages of using the multiple-choice format are: leaving one with the impression that most problems have just one correct or best solution, tempering the role of creative thinking in problem solving, and proving

inadequate for assessing critical thinking dispositions. Third, a source of concern was that the bulk of the questions assessed as testing for critical thinking content were found on the P-ACT+'s Science Reasoning Test. Such a concentration of abilities could send a signal to educators that critical thinking is more important in the field of science than in non-science areas. Fourth, a concern was raised about the subject knowledge requirements needed to answer items judged as calling for a critical thinking ability. An important question to be answered is: Given the desire to assess critical thinking as one of several educational objectives, what amount of subject knowledge requirements should be permitted in tests of general educational achievement? And finally, a brief evaluation was made of the procedure used in this study to determine critical thinking content in the verbal questions of the P-SAT and the P-ACT+.

These considerations, along with the results, suggest possible topics for further research. Although a visual inspection allows one to conclude that the P-SAT and the P-ACT+ are similar in form and content to the SAT and ACT respectively, a replication of this study for the SAT and ACT would provide stronger evidence for drawing such a conclusion. Also, further research is needed on the assessment of both critical thinking content and construct validity in testing, and

the relationship between instruction and testing in critical thinking.

An important reason for undertaking this study was to emphasize some of the issues involved with the assessment of critical thinking on standardized multiple-choice tests like the P-SAT and the P-ACT+. When critical thinking is advanced as an educational objective, and the content of popular standardized multiple-choice exams are influencing curricula and teaching methods, then it becomes important to examine the treatment of critical thinking on these exams. This report discussed several areas for possible improvement in this regard for the P-SAT and the P-ACT+. However, making improvements in the critical thinking content of these exams will not guarantee improvement in critical thinking instruction in the schools. If the perception of critical thinking content that is influencing teaching is algorithmic, piecemeal, and vocabulary oriented, then there runs the risk that more important critical thinking content will be neglected. A lifeless form of critical thinking instruction characterized by the teaching of critical thinking terminology and algorithmic critical thinking operations could result, and the task of teaching students to be creative, independent, critical thinkers could be overlooked. Such a possiblity suggests to educators that school reform bent on teaching critical thinking must consider many

aspects of present-day schooling including other sorts of assessment, not just standardized multiple-choice tests.

A common concern expressed in much of the school reform literature published in the 1980s was that American schools are not facilitating the development of critical thinking in their students; and, given the importance of critical thinking as an educational ideal, improvements must be made. Frequently suggested as one such improvement in that decade of reform was holding schools accountable for measurable educational achievement, most commonly accomplished through the use of multiple-choice standardized tests. This paper suggests, while offering no guarantees, that the accountability movement will serve to enhance the teaching of critical thinking only if teachers, administrators, and curriculum specialists perceive these standardized tests to be testing for critical thinking. Thus the P-SAT and the P-ACT+ exams were analyzed to determine their critical thinking content. Even if these tests are perceived to test for critical thinking, the problems associated with using multiple-choice exams to test for critical thinking speak strongly to the need for alternative assessment methods. The results and the discussion of related issues in this paper suggest that the wedding of standardized multiple-choice exams with the goal

of facilitating the development of critical thinking leads to a potentially unhappy marriage.

References

Berg, R. (1990). Ascertaining the Critical Thinking Content of Tests. Unpublished manuscript, University of Illinois, Illinois Critical Thinking Project, Champaign.

Boyer, E. (1983). High School. New York: Harper & Row.

Ennis, R. H. (1987). A Taxonomy of critical thinking dispositions and abilities. In Baron, J. B., & Sternberg, R. J. (Eds.), Teaching Thinking Skills: Theory and Practice (pp. 9-26). New York: W.H. Freeman.

Hertzka, A. P. & Guilford, J. P. (1955). Logical reasoning. Los Angeles: Instructional Objectives Exchange.

Kotowski, J. (1990). Critical Thinking in the Iowa Tests of Educational Development. Unpublished manuscript, University of Illinois, Illinois Critical Thinking Project, Champaign.

Messick, S. (1989). Meaning and values in test validation: The Science and ethics of assessment. Educational Researcher, 18(2), 5-11.

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. (1983). Washington DC: Author.

Niell, M., & Medina, N. J. (1989). Standardized Testing: Harmful to educational health. Phi Delta Kappan, 70(9), 688-697.

Norris, S.P., & Ennis, R.H. (1989). Evaluating Critical Thinking. Pacific Grove, CA: Midwest Publications.

P-ACT Plus. (1988). Iowa City, IA: The American College Testing Program.

P-ACT+. Form 07A. (1987). Iowa City, IA: The American College Testing Program.

P-SAT. Form T. (1988). Princeton, NJ: Educational Testing Service.

Sizer, T. (1984). Horace's compromise: The Dilemma of the american high school. Boston: Houghton Mifflin.

Walsh, M. (1988) Number of students taking exams grows. Education Week, VIII (4), 23.

Weddle, P. (1985). Fact from opinion. Informal Logic, VII (1),19-26.

Appendix

Evaluation Questionnaire for Determining Critical Thinking Content in Tests

Clarification

Does the question ask the student to . . .

1. Focus on a question other than the test item?
   a. identify or formulate a question?
   b. identify or formulate criteria for judging possible answers?
   c. keep the situation in mind?
2. Analyze the structure of arguments?
   a. identify conclusions?
   b. identify stated reasons?
   c. identify unstated reasons? (implicit used assumptions)
   d. identify unstated needed assumptions.
   e. see and understand the analogy or metaphor?
   f. identify and handle irrelevance?
   g. see the overall structure of an argument?
   h. summarize?
3. Identify argument strategies, and types of statements?
4. Ask and answer questions of clarification and/or challenge?
   for example:
   a. Why?
   b. What is your main point?
   c. What do you mean by "_____"?
   d. What would be an example?
   e. What would not be an example (though close to being one)?
   f. How does that apply to this case (a case that might well appear a counter example)?
   g. What difference does it make?
   h. What are the facts?
   i. Is this what you are saying: _____?
   j. Would you say some more about that?
   k. What are the relevant similarities of your analogy/metaphor?
5. Define terms, and judge definitions in these dimensions?
   a. Form; including:
      1) Synonym.
      2) Classification.
      3) Range.
      4) Equivalent expression.

5) Operational.
6) Example--nonexample.

b. Acts; including:
1) Reporting a meaning- "reported definition"
2) Stipulating a meaning- "stipulative definition"
3) Expressing an opinion on an issue- "positional", including "programmatic" and "persuasive" definition.

c. Content.

6. Know the meaning of "fallacy labels."

Basic Support

7. In judging the credibility of a source, does the question ask the student to use the following criteria:
a. Expertise?
b. Lack of conflict of interest?
c. Agreement among sources?
d. Reputation?
e. Use of established procedures?
f. Known risk to reputation?
g. Ability to give reasons?
h. Careful habits?

8. In evaluating an observation does the question ask the student to consider whether there was:
a. No inferring involved?
b. A short time interval between observation and report?
c. Hearsay?
d. A report based on a record? If so, then was the record:
1) Close to the observation?
2) Made by the observer?
3) Made by the reporter?
4) Was believed by the reporter, either because of a prior belief in its correctness or because of a belief that the observer was habitually correct?
e. Corroboration?
f. Possibility of corroboration?
g. A condition of good access?
h. Competent employment of technology, if technology is useful?

i. Satisfaction by observer (and reporter, if a different person) of credibility criteria? (# 5. above)

Inference

Does the question ask the student to . . .

9. Deduce and/or judge deductions; by using . . .
   a. Class logic?
   b. Conditional logic? Including such things as . . .
      1) Double negation.
      2) Necessary and sufficient conditions.
      3) Appropriate use of logical words such as: "only", "if and only if", "or", "and", "if", "some", "unless", "not", "not both", etc.
10. Induce by generalizing?
   a. Judge the typicality of data: limitation of coverage?
   b. Judge the sampling?
   c. Judge the quantity of data?
   d. Judge the verifiability of the generalization?
11. Induce by inference to best explanation?
   a. Judge the adequacy of one of these types of explanatory conclusions or hypotheses?
      1) Causal claims.
      2) Claims about the beliefs and attitudes of people.
      3) Interpretations of authors' intended meanings.
      4) Historical claims that certain things happened.
      5) Reported definitions.
   b. Investigate by doing such things as . . .
      1) Designing experiments, including planning to control variables?
      2) Seeking evidence and counterevidence?
      3) Seeking other possible explanations?
   c. Employ these criteria? Given reasonable assumptions:
      1) The proposed conclusion would explain the evidence. (essential)
      2) The proposed conclusion is consistent with known facts. (essential)

3) Competitive alternative conclusions are inconsistent. (essential)
4) The proposed conclusion is verifiable. (essential)
5) The proposed conclusion seems plausible. (desirable)

12. Make and judge value judgments by
    a. Considering background facts?
    b. Considering consequences?
    c. Considering prima facie application of acceptable principles?
    d. Considering alternatives?
    e. Weighing and balancing?

Dispositions

Does the question allow the student to demonstrate the ability . . .

13. to be clear about the intended meaning of what is said, written, or otherwise commanded?
14. to determine and maintain focus on the conclusion or question?
15. to take into account the total situation?
16. to seek and offer reasons?
17. to try and be well informed?
18. to look for alternatives?
19. to seek as much precision as the question requires?
20. to try to be reflectively aware of one's own basic beliefs?
21. to be open-minded: consider seriously other points of view than one's own?
22. to withold judgment when the evidence and reasons are insufficient?
23. to take a position (and change a position) when the evidence and reasons are sufficient to do so?
24. to use one's critical thinking abilities?